



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/770,421	02/04/2004	David Llewellyn Rees	01263.001020.1	1942

5514 7590 09/01/2006

FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112

EXAMINER

LERNER, MARTIN

ART UNIT PAPER NUMBER

2626

DATE MAILED: 09/01/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/770,421	Applicant(s) REES, DAVID LLEWELLYN	
	Examiner Martin Lerner	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 17 July 2006 and 24 August 2006.
- 2a) This action is **FINAL**. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 13, 14, 16 to 24, 37, 38, 40 to 48, 50, and 52 is/are pending in the application.
4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 13, 14, 16 to 24, 37, 38, 40 to 48, 50, and 52 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 24 August 2006 is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
a) All b) Some * c) None of:
1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. 09/409,247.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

2. Claims 13, 18, 21, 37, 42, 45, 50, and 52 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.*

Concerning independent claims 13, 37, 50, and 52, *Chigier* discloses an apparatus, method, computer executable process, and computer executable steps, comprising:

“means for receiving the input signal” – an input speech signal 14 is received (column 4, lines 25 to 45: Figure 1);

“means for processing the received signal to generate an energy signal indicative of the local energy within the received signal” – spectral analyzer 12 performs spectral analysis (e.g., computes a short term Fourier transform) on a window of samples to provide a feature vector sequence 16, consisting of a set of parameter coefficients (e.g. cepstral coefficients) characteristic of each speech frame (column 4, lines 46 to 59: Figure 1); cepstral coefficients are “an energy signal indicative of the local energy” because they represent a log energy of a speech signal (Figures 2 and 2A);

“means for determining the likelihood that the boundary is located at each of a plurality of possible locations within the energy signal” – a boundary classifier 54 assigns to each speech frame a probability (“the likelihood”) that the speech frames correspond to a boundary between two phonemes (column 6, lines 10 to 24: Figures 3 and 3A); word boundaries 44 correspond to a case in which an initial sound 50 is classified as part of background signal 52 (“background noise containing portion”) (column 5, line 64 to column 6, line 9: Figures 2 and 2A);

“means for determining the location of the boundary using the likelihoods determined for each of the possible locations” – if a boundary probability assigned to a speech frame is greater than a first threshold (e.g., 70%), the frame is assumed to be a boundary by a segment generator 56, which generates a network of speech segments (A, B, and C); in operation, boundary classifier classifies boundaries I, II, and III in a speech frame sequence 59; segment generator 56 produces speech segments A, B, and C based on the classified boundaries (column 6, lines 15 to 38: Figures 3 and 3A).

Concerning independent claims 13, 37, 50, and 52, *Chigier* discloses detecting phoneme boundaries for speech recognition, but omits “speech detection means operable to process the received signal and to identify when speech is present in the received signal” and “wherein said likelihood determining means is restricted to determine the likelihoods in the received signal only when said speech detecting means detects speech within the received signal.” However, *Chow et al.* teaches a method and apparatus for detecting end points of speech activity, where a VQ distortion processing block 303 performs sound classification to determine whether a sound

waveform is speech or noise. If VQ distortion processing block 303 determines the sound waveform represents speech, then the sound waveform is passed to the speech recognition stage. On the other hand, if VQ distortion processing block 303 determines that the sound waveform represents noise, then the sound waveform is not permitted to pass to the speech recognition stage. (Column 7, Lines 30 to 45: Figure 3) Thus, *Chow et al.* meets the limitations of a speech detection means that detects when speech is present in a received signal, wherein a sound waveform is only passed when speech is detected. An objective is to produce a speech activity detection system that reduces computation in the recognition system. (Column 2, Lines 57 to 59; Column 6, Lines 65 to 67) It would have been obvious to one having ordinary skill in the art to provide a speech detection means so that a likelihood determining means is restricted to determining likelihoods in a received signal only when a speech detecting means detects speech in the received signal as suggested by *Chow et al.* in an automatic speech recognition method and apparatus of *Chigier* for a purpose of reducing an amount of computation required by a speech recognition method and apparatus.

Concerning claims 18 and 42, *Chigier* discloses spectral analyzer 12 blocks a sampled speech signal into frames by placing a "window" over the samples that preserves the samples in the time interval of interest (column 4, lines 45 to 50: Figure 1A).

Concerning claims 21 and 45, *Chigier* discloses word boundaries 44 correspond to a case in which an initial sound 50 is classified as part of background signal 52 (e.g. when sound 50 is a typical mouth click or pop produced by opening the lips, prior to

Art Unit: 2626

speaking), and boundaries 46, correspond to a case in which an initial sound is classified as part of a word (column 5, line 64 to column 6, line 9: Figures 2 and 2A); implicitly, at least a boundary at a beginning of a speech portion is detected.

3. Claims 14, 22, 38, and 46 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* as applied to claims 13 and 37 above, and further in view of *Cohrs et al.*

Concerning claims 14 and 38, *Chigier* discloses checking boundary probability classifications of one or more frames from either side of frame N (column 6, line 65 to column 7, line 1), but omits determining a boundary location by comparing with a model representative of energy in background noise and a model representative of energy in speech, and combining results of the comparisons to determine a likelihood for a current location. However, *Cohrs et al.* teaches computation of a similarity measure between stored references and parameters extracted from an utterance using hidden Markov models (HMMs). Hypothesizer 43 makes two types of hypotheses. The first type of hypothesis (referred to as a "background hypothesis") assumes that the feature vector sequence includes only background. The second type of hypothesis (referred to as a "phrase hypothesis") assumes that the feature sequence includes a command word. (Column 4, Line 59 to Column 5, Line 20: Figure 2) *Cohrs et al.* states there is an advantage in using models instead of thresholds for spotting command words by avoiding problems associated with false alarm rates for certain users. (Column 1, Lines 31 to 63) It would have been obvious to one having ordinary skill in the art to determine

Art Unit: 2626

boundaries by comparing to models of background noise and speech as taught by *Cohrs et al.* in the method and apparatus for boundary probability assignment of *Chigier* for the purpose of avoiding problems associated with using thresholds.

Concerning claims 22 and 46, *Cohrs et al.* teaches hidden Markov models (HMMs) (column 4, lines 1 to 5), which are statistical models, implicitly.

4. Claims 16, 17, 19, 20, 40, 41, 43, and 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* as applied to claims 13 and 37 above, and further in view of *Lennig et al.*

Concerning claims 16, 17, 40, and 41, *Chigier* omits filtering an energy signal to remove energy variations having a frequency below a predetermined frequency, where the filter is operable to filter out energy variations below 1 Hz. However, *Lennig et al.* teaches detecting word endpoints, where filter means 12 comprises a filter bank of twenty triangular filters spanning a range of about 100 Hz to about 4000 Hz. Weights W_{ij} for filter channels j are set so that $W_{ij} = 0$ for frequencies f_j below 100 Hz. (Column 3, Lines 4 to 40; Figure 1; Table 1: Filter No. 1) Thus, all energy variations at frequencies in the range between 0 Hz and 100 Hz are removed, including those energy variations at frequencies below 1 Hz. *Lennig et al.* suggests an advantage of reducing an error rate for speech recognition. (Column 1, Lines 19 to 26) It would have been obvious to one having ordinary skill in the art to filter an energy signal to remove energy variations having a frequency below a predetermined frequency as taught by *Lennig et al.* in the

Art Unit: 2626

method and apparatus of boundary probability assignment of *Chigier* for the purpose of reducing an error rate for speech recognition.

Concerning claims 19, 20, 43, and 44, *Chigier* discloses speech samples (column 4, lines 60 to 66), and assigning boundary probabilities based on log energy (column 6, lines 10 to 24: Figures 2, 2A, and 3).

5. Claims 23 and 47 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* and *Cohrs et al.* as applied to claims 13, 14, 22, 37, 38, and 46 above, and further in view of *Abut et al.*

Cohrs et al. discloses hidden Markov models (HMMs), but omits models based on Laplacian statistics. However, *Abut et al.* discloses speech probability models based on Laplacian speech statistics. (II. Speech Statistics: Page 226) It is suggested that Laplacian statistics have lower and upper bounds suitable for speech probability models. (Page 227) It would have been obvious to one having ordinary skill in the art to utilize models based upon Laplacian statistics as suggested by *Abut et al.* in the method and apparatus for boundary probability assignment of *Chigier* in order to obtain suitable speech probability models.

Claims 24 and 48 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* and *Cohrs et al.* as applied to claims 13, 14, 22, 37, 38, and 46 above, and further in view of *Erell et al.*

Cohrs et al. discloses hidden Markov models (HMMs), but does not expressly state that a speech model is an auto-regressive model. However, *Erell et al.* teaches a speech recognition system where the acoustic features are extracted to form a feature vector, and where the features are the coefficients of an autoregressive model. *Erell et al.* states that these are the most commonly used features, including linear prediction coefficients, cepstrum coefficients, bank of filter energies etc., to reflect vocal tract characteristics. (Column 1, Lines 37 to 45) It would have been obvious to one of ordinary skill in the art to use an auto-regressive model in the method and apparatus for boundary probability assignment of *Chigier* because *Erell et al.* suggests that an auto-regressive model is the most commonly employed method of deriving speech features.

Response to Arguments

6. Applicants' arguments filed 17 July 2006 have been considered but are moot in view of the new grounds of rejection.

Applicant's arguments directed to the rejection of claims 13, 14, 16 to 24, 37, 38, 40 to 48, 50, and 52 under 35 U.S.C. §112, 1st Paragraph, as failing to meet the written description requirement, are persuasive.

Applicants argue that the claimed apparatus and method involves a two-stage speech detector, in which a second stage can be thought of as a fine tuning or more precise determination of speech detection achieved by a first stage. Applicants say that *Chigier* identifies boundaries using only a single stage, and that combining teachings of two single-stage speech detectors would be understood to be a result of hindsight.

Art Unit: 2626

Applicants maintain that a two-stage detector is more than a sum of two single-stage detectors. These arguments are not persuasive.

Chow et al. provides a motivation to incorporate a preliminary determination of whether speech is present in a received sound signal so that further processing is performed only when speech is present. The objective is to reduce the amount of further processing required. *Chigier* is directed to detecting boundaries between phonemes during speech recognition. Clearly, one having ordinary skill in the art would know that if speech is not present, there is no need to perform speech recognition. More specifically, speech recognition requires an identification of phonemes, and identification of phonemes is improved by detecting phoneme boundaries. However, if speech is not present, there will be no phonemes. If a received signal contains only background noise, there are no phonemes in the received signal, and thus, boundaries between phonemes cannot be detected. Computational saving are thereby achieved by not requiring a speech recognition method and apparatus to even attempt to recognize phoneme and phoneme boundaries when there is only noise and no speech present in a received signal.

Applicants' argument that their method and apparatus for boundary detection achieves more than two single-stage speech detectors is not persuasive. The fact that Applicants may have recognized another advantage which would flow naturally from following the suggestion of the prior art cannot be the basis for patentability when the differences would otherwise be obvious. See *Ex parte Obiaya*, 227 USPQ 58, 60 (Bd. Pat. App. & Inter. 1985). *Chow et al.* provides a motivation for combination with *Chigier*

Art Unit: 2626

to reduce computational requirements. A combination would not in any way change the operation of the *Chigier* because one skilled in the art would readily understand that phoneme detection in speech recognition would not be necessary when a received signal contains only background noise. Accordingly, it is maintained that no hindsight is involved in formulating the combination.

Therefore, the rejections of claims 13, 18, 21, 37, 42, 45, 50, and 52 under 35 U.S.C. §103(a) as being unpatentable over *Chigier* in view of *Chow et al.*; of claims 14, 22, 38, and 46 under 35 U.S.C. §103(a) as being unpatentable over *Chigier* in view of *Chow et al.*, and further in view of *Cohrs et al.*; of claims 16, 17, 19, 20, 40, 41, 43, and 44 under 35 U.S.C. §103(a) as being unpatentable over *Chigier* in view of *Chow et al.*, and further in view of *Lennig et al.*; of claims 23 and 47 under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* and *Cohrs et al.*, and further in view of *Abut et al.*; and of claims 24 and 48 under 35 U.S.C. 103(a) as being unpatentable over *Chigier* in view of *Chow et al.* and *Cohrs et al.*, and further in view of *Erell et al.*, are proper.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to Applicants' disclosure.

Gerson et al. discloses a speech activity detector 114 providing a signal 116 as to whether a device controller 120 should perform speech recognition with speech recognizer 110. (Column 4, Lines 2 to 11; Column 5, Lines 60 to 67: Figure 1)

Art Unit: 2626

8. Applicants' amendment necessitated the new grounds of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for

Art Unit: 2626

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

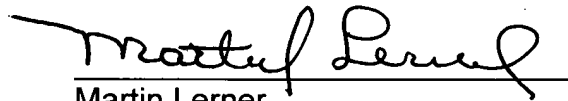
For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free).

ML

8/28/06

A handwritten signature in black ink, appearing to read "Martin Lerner", written over a horizontal line.

Martin Lerner

Examiner

Group Art Unit 2626